**Claims**

What is claimed is:

1. A method of providing speaker recognition, said method comprising the steps of:

providing a model corresponding to a target speaker, the model being resolved into at least one frame and at least one level of phonetic detail;

receiving an identity claim;

ascertaining whether the identity claim corresponds to the target speaker model;

said ascertaining step comprising the steps of:

determining, for each frame and each level of phonetic detail of the target speaker model, a non-interpolated likelihood value; and

resolving the at least one likelihood value to obtain a likelihood score.

2. The method according to Claim 1, wherein, for each frame and each level of phonetic detail, the non-interpolated likelihood value is a maximum likelihood value.

3. The method according to Claim 2, wherein said step of resolving the at least one likelihood value comprises averaging the at least one likelihood value.

4. The method according to Claim 3, wherein the likelihood value is determined via the following general equation:

$$S(U \mid M) = \frac{1}{T}\sum_{i=1}^{L}\sum_{t=1}^{T} b_{i,j(i,t)} \bullet P(u_t \mid M\{i, j(i,t)\}) \ ;$$

wherein b_{i,j(i,t)} corresponds to grain-specific weights that satisfy

$$\sum_{i=1}^{L}\sum_{j=1}^{K(i)} b_{ij} = 1 \ ;$$

and further wherein:

$S$ is the likelihood score;

$U$ is a test utterance, comprising $T$ frames $u_1 \ldots, u_T$;

$M(i,j)$ is a speaker model, with $1 \le i \le L$ levels of detail and with $1 \le j \le K(i)$ units on the $i$-th level; and

$P(u_t|M(i,j))$ is the probability that a frame $u_t$ corresponds to a speaker model unit $j$ on the $i$-th level of phonetic detail of the speaker model.

5. The method according to Claim 4, wherein the likelihood score is determined by the following equation:

$$S(U \mid M) = \frac{1}{T} \sum_{t=1}^{T} \max_{1 \le i \le L, 1 \le j \le K(i)} P(u_t \mid M(i,j)) \ .$$

6. The method according to Claim 1, wherein the at least one level of phonetic detail comprises at least one of the following: a global level; a phonemic level and a sub-phonemic level.

7. The method according to Claim 6, wherein the at least one level of phonetic detail comprises all of the following three levels: a global level; a phonemic level and a sub-phonemic level.

8. The method according to Claim 7, wherein said step of providing a model corresponding to a target speaker comprises creating said target speaker model on the basis of training utterances and providing labeling information for each frame.

9. The method according to Claim 1, wherein said ascertaining step further comprises accepting or rejecting the identity claim.

10. The method according to Claim 9, wherein said step of accepting or rejecting comprises comparing a quantity based on the likelihood score to a predetermined threshold value.

11. The method according to Claim 10, further comprising the steps of:

providing at least one model corresponding to at least one background speaker; and

determining the quantity based on the likelihood score via employing the at least one background speaker model.

12. The method according to Claim 11, wherein said step of determining the quantity based on the likelihood comprises determining a log-likelihood ratio based on the likelihood score.

13. The method according to Claim 12, wherein the log-likelihood ratio is determined by the following equation:

$$L = S(U \mid M) - \frac{1}{C}\sum_{i=1}^{C} S(U \mid BG_i);$$

wherein:

$L$ is the log-likelihood ratio;

$S$ is the likelihood score;

$M$ denotes the target speaker model; and

$BG_i$ denotes the $i$-th background model.

14. An apparatus for of providing speaker recognition, said apparatus comprising:

a target speaker model generator for generating a model corresponding to a target speaker, the model being resolved into at least one frame and at least one level of phonetic detail;

a receiving arrangement for receiving an identity claim;

a decision arrangement for ascertaining whether the identity claim corresponds to the target speaker model;

said decision arrangement being adapted to:

determine, for each frame and each level of phonetic detail of the target speaker model, a non-interpolated likelihood value; and

resolve the at least one likelihood value to obtain a likelihood score.

15. The apparatus according to Claim 14, wherein, for each frame and each level of phonetic detail, the non-interpolated likelihood value is a maximum likelihood value.

16. The apparatus according to Claim 15, wherein said decision arrangement is adapted to resolve the at least one likelihood value via averaging the at least one likelihood value.

17. The apparatus according to Claim 16, wherein the likelihood value is determined via the following general equation:

$$S(U \mid M) = \frac{1}{T} \sum_{i=1}^{L} \sum_{t=1}^{T} b_{i,j(i,t)} \cdot P(u_t \mid M\{i, j(i,t)\}) \ ;$$

wherein b_{i,j(i,t)} corresponds to grain-specific weights that satisfy

$$\sum_{i=1}^{L} \sum_{j=1}^{K(i)} b_{ij} = 1$$

and further wherein:

$S$ is the likelihood score;

$U$ is a test utterance, comprising $T$ frames $u_1, \ldots, u_T$;

$M(i,j)$ is a speaker model, with $1 \leq i \leq L$ levels of detail and with $1 \leq j \leq K(i)$ units on the $i$-th level; and

$P(u_t|M(i,j))$ is the probability that a frame $u_t$ corresponds to a speaker model unit $j$ on the $i$-th level of phonetic detail of the speaker model.

18. The apparatus according to Claim 17, wherein the likelihood score is determined by the following equation:

$$S(U|M) = \frac{1}{T}\sum_{t=1}^{T} \max_{1 \leq i \leq L, 1 \leq j \leq K(i)} P(u_t|M(i,j)) \ .$$

19. The apparatus according to Claim 14, wherein the at least one level of phonetic detail comprises at least one of the following: a global level; a phonemic level and a sub-phonemic level.

20. The apparatus according to Claim 19, wherein the at least one level of phonetic detail comprises all of the following three levels: a global level; a phonemic level and a sub-phonemic level.

21. The apparatus according to Claim 20, wherein said target speaker model generator is adapted to generate said target speaker model on the basis of training utterances and providing labeling information for each frame.

22. The apparatus according to Claim 14, wherein said decision arrangement is further adapted to accept or reject the identity claim.

23. The apparatus according to Claim 22, wherein said decision arrangement is adapted to accept or reject the identity claim via comparing a quantity based on the likelihood score to a predetermined threshold value.

24. The apparatus according to Claim 23, further comprising:

a background speaker model generator for providing at least one model corresponding to at least one background speaker;

said decision arrangement being adapted to determine the quantity based on the likelihood score via employing the at least one background speaker model.

25. The apparatus according to Claim 24, wherein said decision arrangement is adapted to determine the quantity based on the likelihood via determining a log-likelihood ratio based on the likelihood score.

26. The apparatus according to Claim 25, wherein the log-likelihood ratio is determined by the following equation:

$$L = S(U \mid M) - \frac{1}{C}\sum_{i=1}^{C} S(U \mid BG_i);$$

wherein:

$L$ is the log-likelihood ratio;

$S$ is the likelihood score;

$M$ denotes the target speaker model; and

$BG_i$ denotes the $i$-th background model.

27. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for providing speaker recognition, said method comprising the steps of:

providing a model corresponding to a target speaker, the model being resolved into at least one frame and at least one level of phonetic detail;

receiving an identity claim;

ascertaining whether the identity claim corresponds to the target speaker model;

said ascertaining step comprising the steps of:

determining, for each frame and each level of phonetic detail of the target speaker model, a non-interpolated likelihood value; and

resolving the at least one likelihood value to obtain a likelihood score.